ENSIAME

Optimisation:

De l'estimation paramétrique à l'apprentissage, une ballade entre théorie et pratique

S. Delprat

Chapitre 1 – Concepts mathématiques

Préambule

Qu’est ce qui est abordé dans cet ensemble de cours/TD ?

- Présentation de différents algorithmes d’optimisation
- Les formules mathématiques sous-jacentes

Qu’est ce qui n’est pas abordé dans cet ensemble de cours/TD ?

- Les preuves de convergence
- Les aspects liés à la mise en œuvre numérique : factorisation, inversion de matrice, etc.

1) Problématiques abordées en cours
2) Ecriture d’un problème d’optimisation
3) Minimisation non contrainte
4) Minimisation avec contraintes
5) Convexité

Introduction

- Une partie significative du travail d'ingénieur consiste à optimiser, améliorer un procédé.

Trouver la meilleure position

Réduire la consommation d'énergie

Améliorer les performances d'un système

Minimiser les pertes

Diminuer le temps de fabrication

Chercher le meilleur rapport qualité/prix

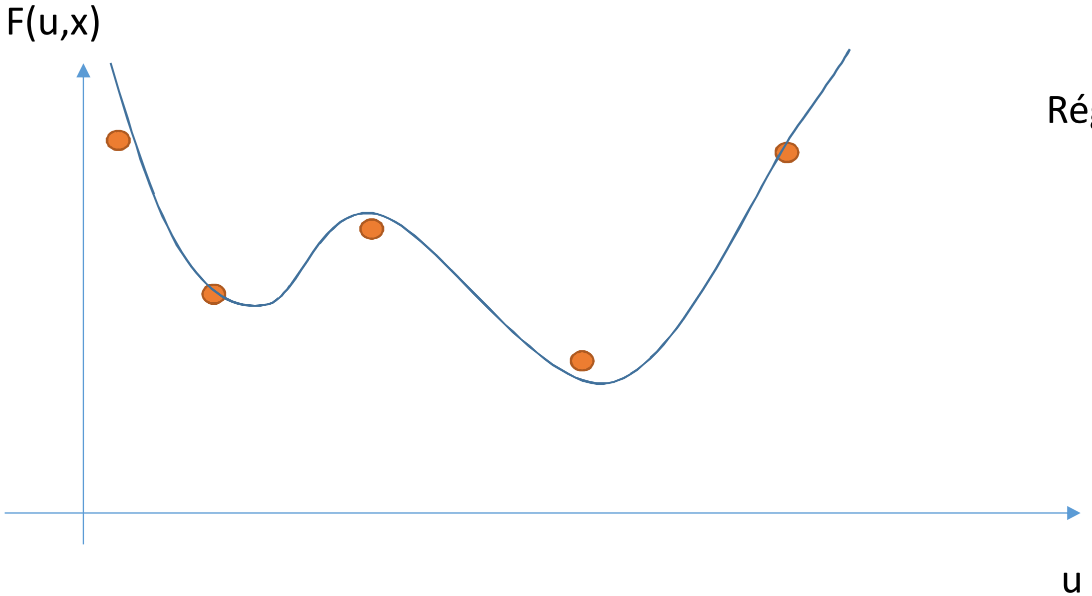
Lorsqu'un critère est défini, en général, il est possible de formuler un problème d'optimisation

Introduction

- Une autre classe de problèmes consiste à trouver les paramètres d'une fonction de manière à obtenir un résultat

Estimation paramétrique

$$y = f(u)$$



Régression polynomiale:

Trouver les coefficients p_i qui minimisent l'erreur de modèle

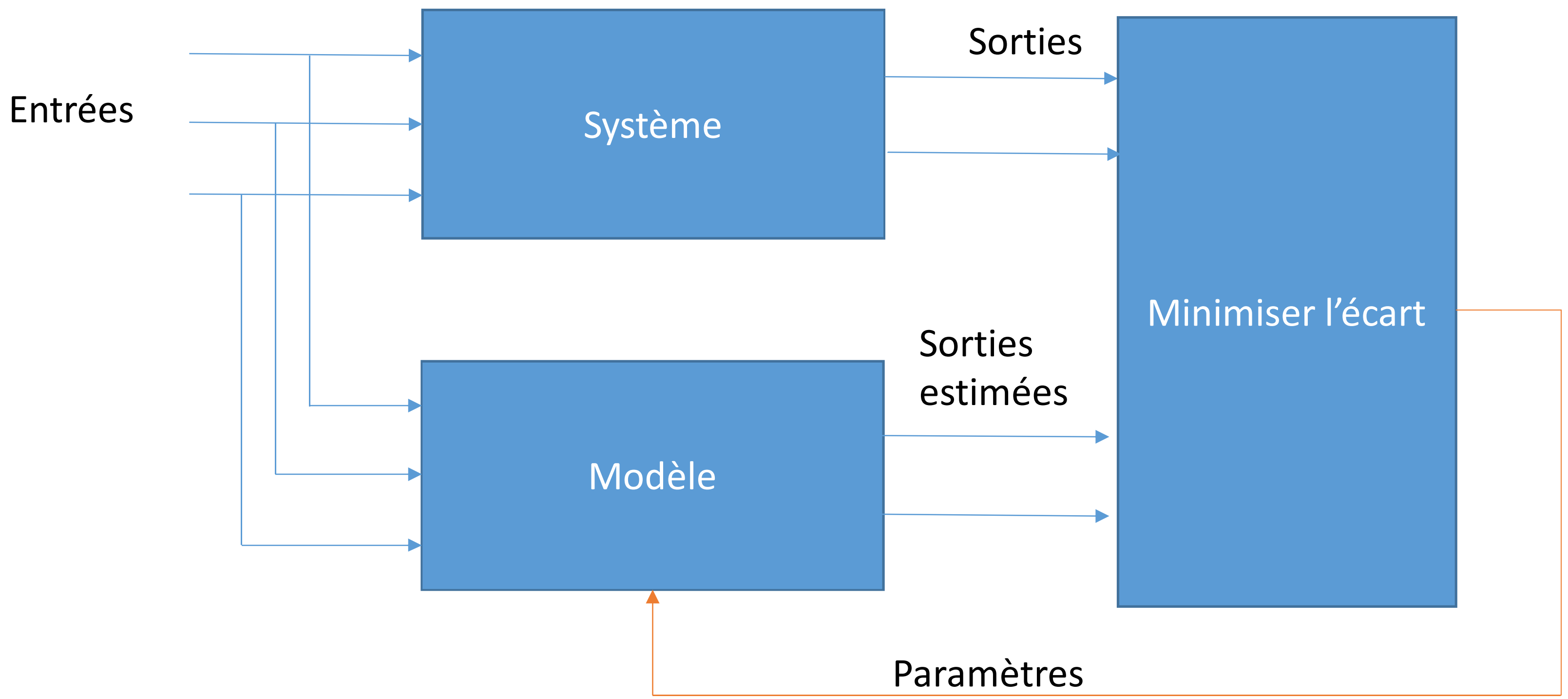
Modèle: $$\hat{y} = \sum_{i=0}^{n-1} p_i u^i$$

Erreur de modèle $$\varepsilon = \sum_{j=0}^{m-1} \left(y_j - \hat{y}_j \right)^2$$

Introduction

- Une autre classe de problème consiste à trouver les paramètres d’une fonction de manière à obtenir un résultat

Estimation paramétrique



- Comment faire si on n'a peu ou pas d'information sur la fonction f ?

$\Rightarrow$Utiliser un modèle « suffisamment flexible » pour s'adapter à n'importe quelle fonction f *(avec quelques restrictions quand même)*.

$\Rightarrow$Propriété d'approximation universelle

Une approche : les réseaux de neurones…

<u>Applications :</u>

Reconnaissance d'images : classification, interprétation de scènes, etc.
Commande de systèmes
Traduction de texte
Compréhension du langage naturel, etc.

En 2017 ces fonctionnalités requièrent l'optimisation de quelques millions de paramètres
$\Rightarrow$ Il faut beaucoup de données d'entrées « big data »
$\Rightarrow$ Temps d'exécution de l'algorithme d'optimisation : quelques semaines sur des super calculateurs
Mais le modèle résultat est exploitable en temps reel

1) Problématiques abordées en cours
2) Ecriture d'un problème d'optimisation
3) Minimisation non contrainte
4) Minimisation avec contraintes
5) Convexité

Introduction – L'optimisation pourquoi faire?

- Variables: Ce que l'on veut trouver

Scalaire ou vecteur	$x \in \mathbb{R}$	1 variable réelle
Binaire, entier ou Réel	$x \in \mathbb{R}^n$	n variables réelles
	$x \in \mathbb{R} \times \{0,1\}^2$	1 variable réelle + 2 variables binaires
	etc.	

- Critère: Ce que l'on veut minimiser/maximiser

C'est une fonction qui doit pouvoir être évaluée.

Elle peut être :

- Linéaire / non linéaire
- Différentiable / non différentiable
- Définie sur un support bornée ou non
- Convexe ou non
- Non linéaire, non différentiable:

$J = 2x + 2$	Linéaire , différentiable
$J = x^2$	Non linéaire , différentiable
$J = \mathrm{mod}(x,4)$	Non linéaire , non différentiable

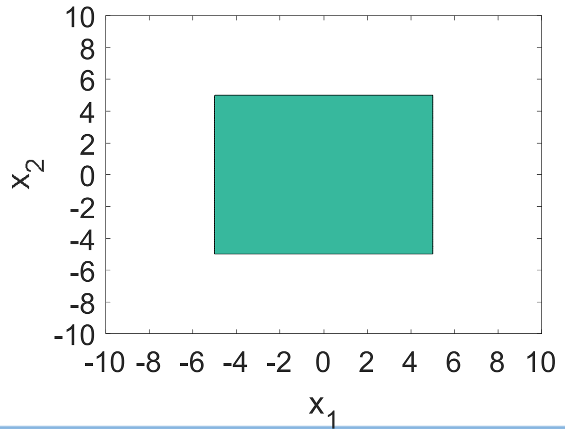
Simulation simulink, CFD, etc.

Introduction – L'optimisation pourquoi faire?

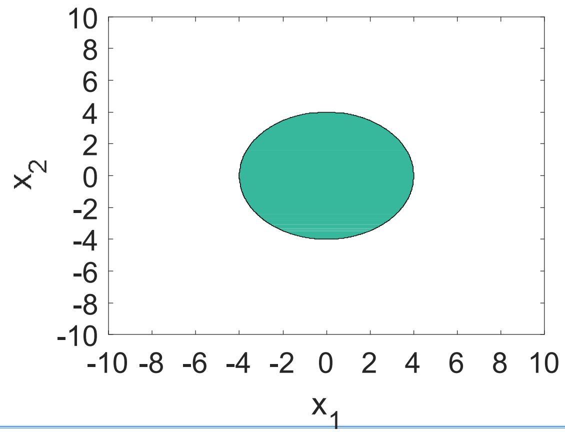
- Contraintes: Dans quel espace doit-on chercher la solution

Egalités	$2x_1 + 3x_2 = 0$	$x_1{}^2 + x_2{}^2 = 2$
Inégalités	$2x_1 + 3x_2 > 0$	$x_1{}^2 + x_2{}^2 = 2$

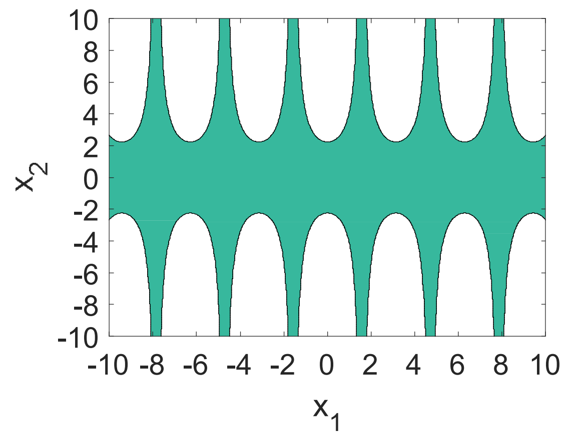
$$\max\left(|x_1|,|x_2|\right) < 5$$



$$\max\left(x_1{}^2 + x_2{}^2\right) < 16$$



$$\left(\cos\left(x_1{}^2\right)x_2\right)^2 < 5$$



Introduction – L'optimisation pourquoi faire?

- Contraintes: Dans quel espace doit on chercher la solution

Contraintes linéaires:

- Egalité : $Ax = b$

  Ce type de contrainte définit un hyper-plan (droite, plan, etc.)

- Inégalité : $Ax > b$

  Définit l'espace au dessus ou dessous de l'hyper plan

Un peu de vocabulaire pour commencer

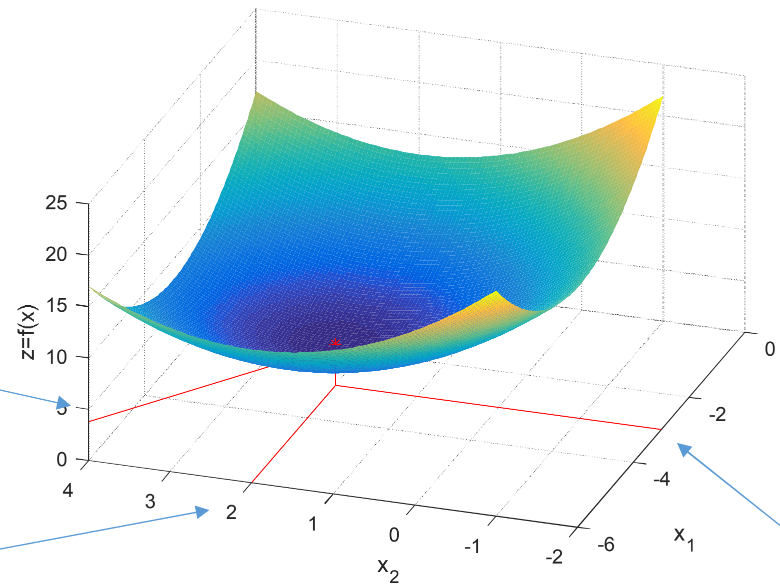
On cherche à minimiser un critère f(x). L'objectif est donc de trouver :

- Le minimum/maximum de f $\quad J=\min\limits_{dom(f)} f(x) \qquad J=\max\limits_{dom(f)} f(x)$

- La valeur x* qui minimise/maximise f : $\quad x^*=\arg\min\limits_{dom(f)} f(x) \qquad x^*=\arg\max\limits_{dom(f)} f(x)$

$$f(x)=(x+3)^2+(y-2)^2+4$$

$$J=\min_{dom(f)} f(x)=4$$



$$x^*=\arg\min_{dom(f)} f(x)=(-3,2)^T$$

Un peu de vocabulaire pour commencer

Soit une fonction $f : S \to \mathbb{R}$ définie sur un sous ensemble $S \subseteq \mathbb{R}^n$

<u>Minimum global:</u>

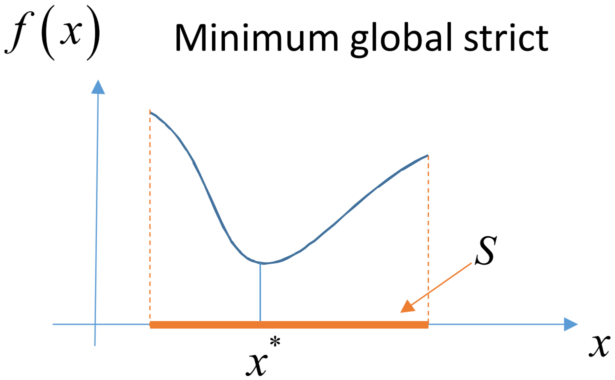
$x^* \in S$ est un minimum global de f sur S si $f\left(x^*\right) \leq f\left(x\right) \quad \forall x \in S$

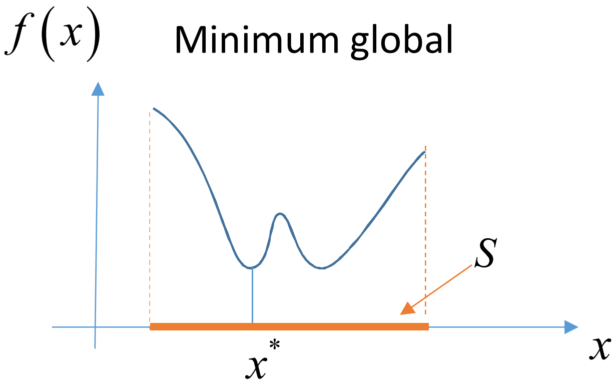
$x^* \in S$ est un minimum global strict de f sur S si $f\left(x^*\right) < f\left(x\right) \quad \forall x \in S / x^*$

<u>Maximum global:</u>

$x^* \in S$ est un maximum global de f sur S si $f\left(x^*\right) \geq f\left(x\right) \quad \forall x \in S$

$x^* \in S$ est un maximum global strict de f sur S si $f\left(x^*\right) > f\left(x\right) \quad \forall x \in S / x^*$





Un peu de vocabulaire pour commencer

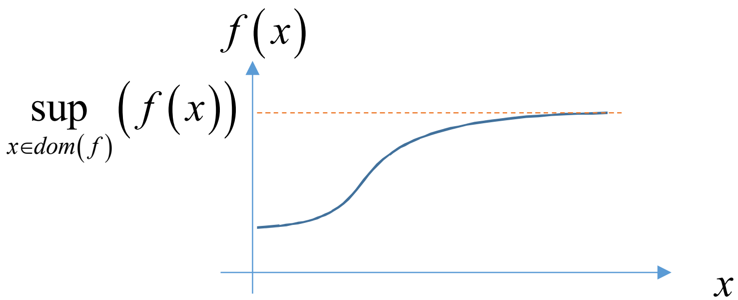
Attention :

$\max_{x \in S} f(x)$ est unique

Mais la fonction f peu admettre plusieurs minimum globaux:

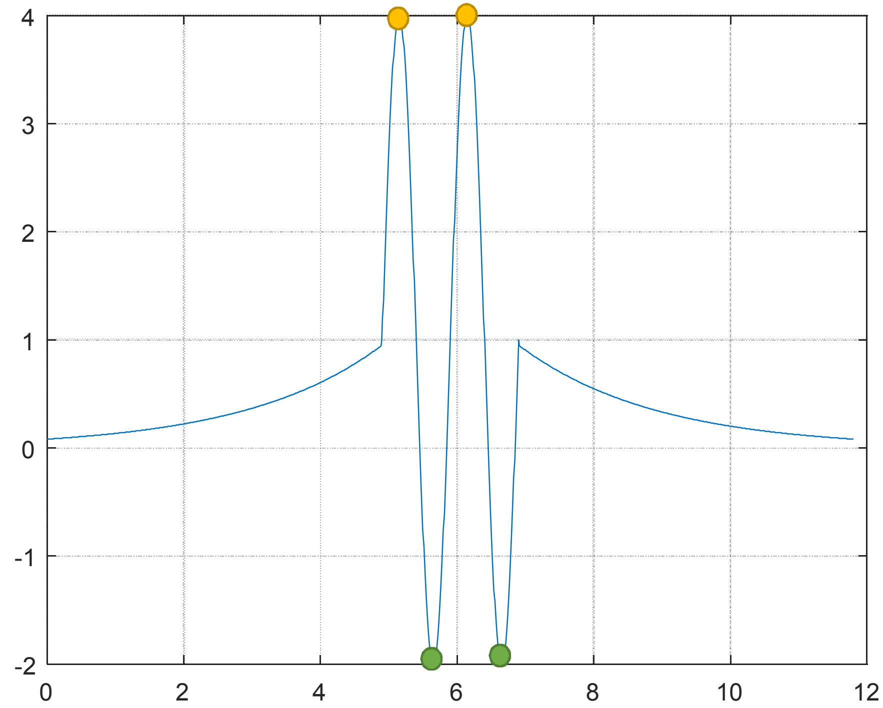
$\arg\max_{x \in S} f(x)$ n’est pas unique !!

NB: Dans le cadre de ce cours, on supposera que la fonction f atteint son maximum/minium : on ne traite pas le cas des asymptotes



$\max_{x \in S} f(x) = 4$ $\arg\max_{x \in S} f(x) = \{5.1, 6.1\}$

Plusieurs maxima globaux



Plusieurs minimum globaux

Un peu de vocabulaire pour commencer

Soit une fonction $f : S \rightarrow \mathbb{R}$ définie sur un sous ensemble $S \subseteq \mathbb{R}^n$

<u>Minimum local:</u>

$x^* \in S$ est un minimum *local* de f sur S si il existe r>0 tel que $f\left(x^*\right) \leq f\left(x\right) \qquad \forall x \in S \cap \overbrace{B\left(x^*, r\right)}^{\text{Boule de centre } x^* \text{ de rayon } r}$

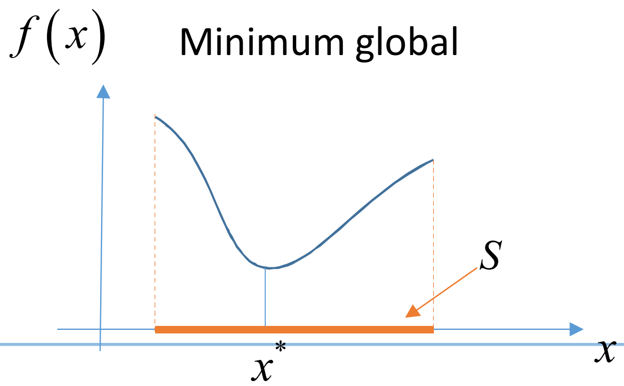
$x^* \in S$ est un minimum *local strict* de f sur S si il existe r>0 tel que $f\left(x^*\right) < f\left(x\right) \qquad \forall x \in S \cap B\left(x^*, r\right) / x^*$

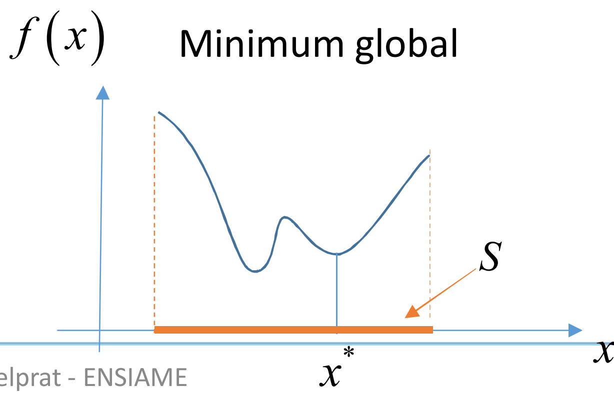
<u>Maximum local:</u>

$x^* \in S$ est un maximum *local* de f sur S si il existe r>0 tel que $f\left(x^*\right) \geq f\left(x\right) \qquad \forall x \in S \cap B\left(x^*, r\right)$

$x^* \in S$ est un maximum *local strict* de f sur S si il existe r>0 tel que $f\left(x^*\right) > f\left(x\right) \qquad \forall x \in S \cap B\left(x^*, r\right) / x^*$

NB: un minimum (resp. maximum) global est nécessairement un minimum (resp. maximum) local





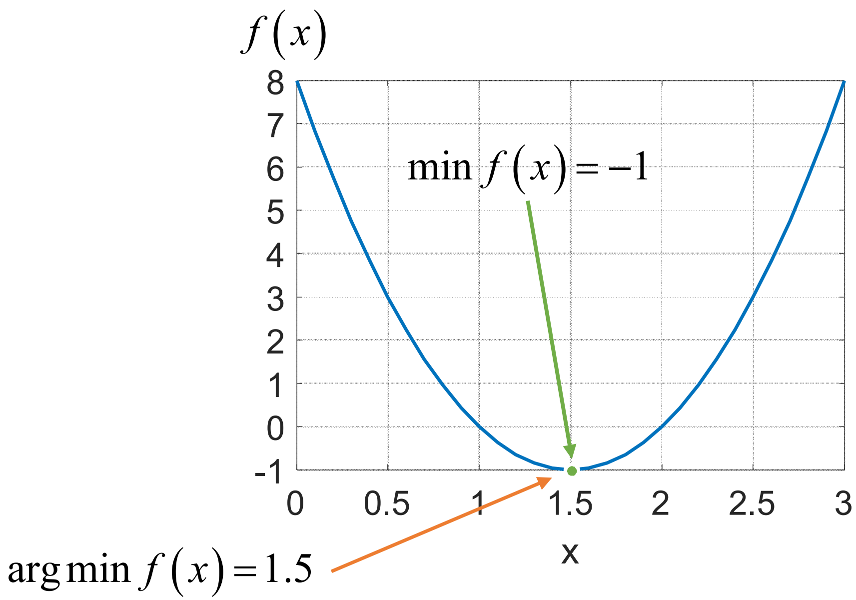
Propriétés

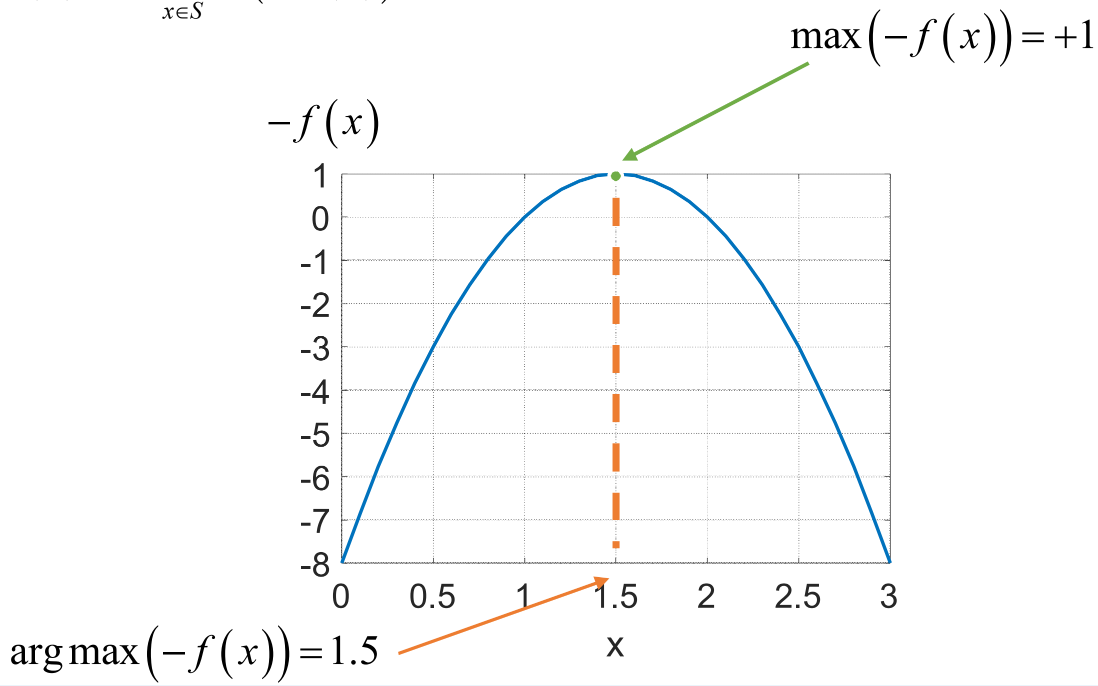
Soit une fonction $f : S \to \mathbb{R}$ définie sur un sous ensemble $S \subseteq \mathbb{R}^n$

On peut facilement passer d'un problème de minimisation à un problème de maximisation:

$$\min_{x \in S} f(x) = -\max_{x \in S}\left(-f(x)\right) \qquad \arg\min_{x \in S} f(x) = \arg\max_{x \in S}\left(-f(x)\right)$$

Et l'inverse: $\max_{x \in S} f(x) = -\min_{x \in S}\left(-f(x)\right) \qquad \arg\max_{x \in S} f(x) = \arg\min_{x \in S}\left(-f(x)\right)$





1) Problématiques abordées en cours
2) Ecriture d’un problème d’optimisation
3) Minimisation non contrainte
4) Minimisation avec contraintes
5) Convexité

Gradient

Soit une fonction $f:\mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^1$

Gradient d'une fonction:

Généralisation de la notion de dérivée aux fonctions de plusieurs variables.

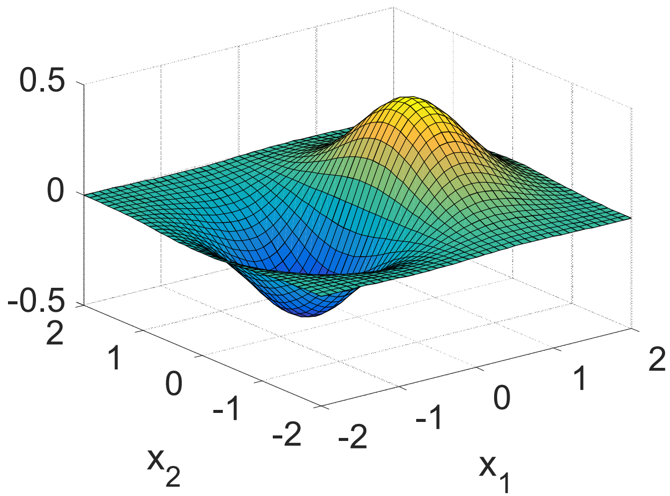
En chaque point de l'espace, le gradient est un vecteur qui pointe dans la direction de la plus grande « croissance » de f.

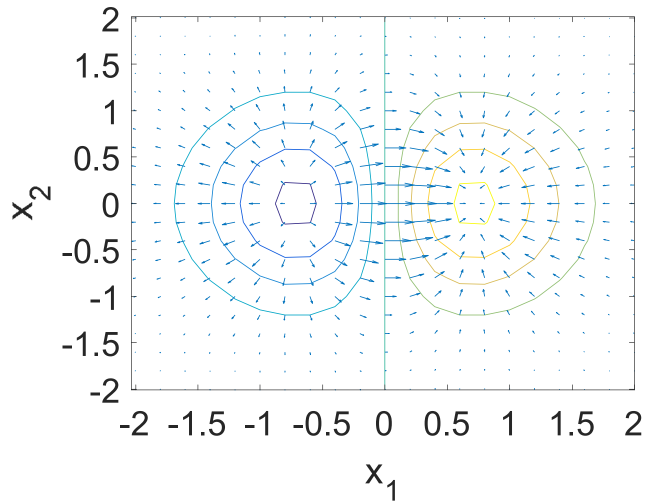
$$\vec{\nabla} f(x) = \nabla f(x) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{pmatrix}$$

Operateur « Nabla »

$$f(x) = x_1 e^{-x_1{}^2 - x_2{}^2}$$

$$\nabla f(x) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1} = -(2x_1 - 1)e^{-x_1{}^2 - x_2{}^2} \\ \dfrac{\partial f}{\partial x_2} = -2x_1 x_2 e^{-x_1{}^2 - x_2{}^2} \end{pmatrix}$$





Gradient

<u>Utilisation de la toolbox symbolique de Matlab:</u>

```
clear all
close all;
clc;
```

« Nettoyage » écran, mémoire

```
% Calcul symbolique
syms x1 x2 real
```

Création des objets symboliques

```
f = x1.*exp(-x1^2-x2^2);
g = gradient(f, [x1, x2]);
```

Définition de la fonction
Calcul gradient (symbolique)

```
% Création de fonctions
g1=symfun(g(1),[x1,x2]);
g2=symfun(g(2),[x1,x2]);
f=symfun(f,[x1,x2]);
```

Création de fonctions à partir des expressions symboliques

```
% Evaluation numérique des fonctions sur une grille
[X1, X2] = meshgrid(-2:.2:2,-2:.1.5:2);
F = f(X1,X2);
G1 = g1(X1,X2);
G2 = g2(X1,X2);
```

Définition d'une « grille » carrée + évaluation des fonctions et du gradient sur la grille

```
% Conversion symbolique => double
F=double(F);
G1=double(G1);
G2=double(G2);
```

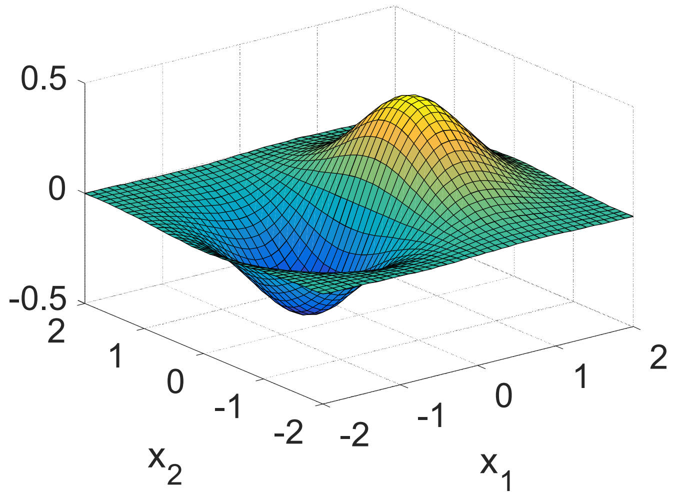
Conversion du résultat (nombre « symbolique ») en double

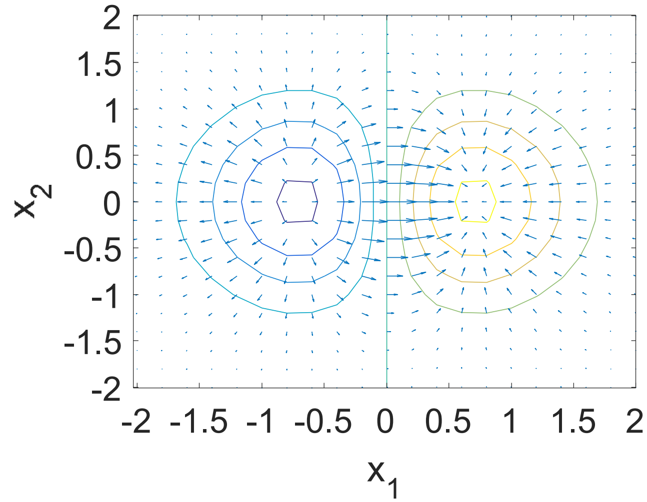
```
>> f
f(x1, x2) =
x1*exp(- x1^2 - x2^2)
>> g
g =
 exp(- x1^2 - x2^2) - 2*x1^2*exp(- x1^2 - x2^2)
                -2*x1*x2*exp(- x1^2 - x2^2)
```

Gradient

<u>Utilisation de la toolbox symbolique de Matlab:</u>

```
% Affichages
figure;
surf(X1,X2,F);
hold on;
xlabel('x_1');
xlabel('x_2');

figure;
contour(X1,X2,F);
hold on;
quiver(X1, X2, G1, G2);
xlabel('x_1');
xlabel('x_2');
```





Savoir-faire n°1 :

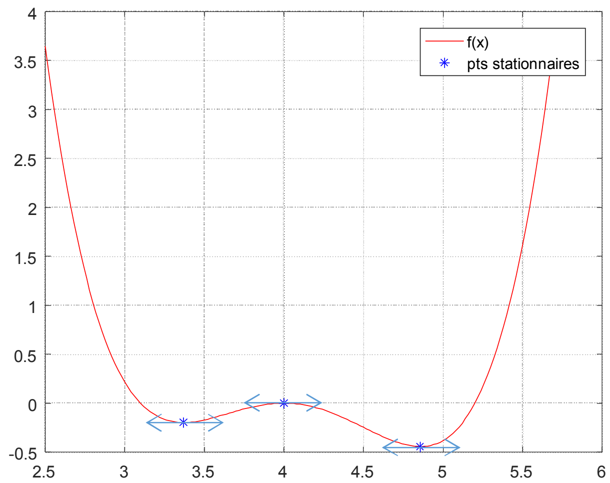
- Définir une fonction symbolique
- Calculer et afficher le gradient d’une fonction de deux variables réelles

Conditions nécessaires d'optimalité

Soit une fonction $f : \mathbb{R} \to \mathbb{R}$ de classe $\mathcal{C}^1$
Soit x* un minimum ou maximum **<u>local</u>**:

Condition nécessaire au premier ordre: $\nabla f\left(x^*\right) = 0$

Les valeurs x* telles que $\nabla f\left(x^*\right) = 0$ sont appelées *points stationnaires* de f.



Les minimum globaux sont aussi des minimum locaux

Les minimums locaux sont aussi des points stationnaires

$\Rightarrow$ La condition au premier ordre ne permet pas immédiatement de trouver le minimum global
$\Rightarrow$ il faut faire le tri parmi les valeurs possibles

Calcul de points stationnaires

Utilisation de la toolbox symbolique de Matlab:

```
clear all
close all;
clc;
```

« Nettoyage » écran, mémoire

```
syms x real

% Fonction à étudier
f=x^4 - (163*x^3)/10 + (2463*x^2)/25 - (6544*x)/25 + 6448/25;
```

Définition de la fonction à étudier

```
% Calcul de la dérivée
dfdx=diff(f,x)
```

Calcul symbolique de $\frac{df}{dx}$

```
% Rechche des points stationnaires
x_stat=solve(dfdx,x)
```

Tentative de résolution de $\frac{df(x)}{dx} = 0$

```
% Tracé de la fonction
X = (2.5:.02:5.7);
f=symfun(f,x);
F = f(X);
F = double(F);
```

Calculs de quelques points $\left(x_i, f(x_i,)\right)$

```
figure;
plot(X,F,'r');
hold on;
plot(x_stat,f(x_stat),'b*');
grid on;
legend('f(x)','pts stationnaires');
```

Affichage

Savoir-faire n°2 : calculer les points stationnaires d'une fonction

Conditions nécessaires d'optimalité

Une matrice $M \in \mathbb{R}^{n\times n}$ est dite *définie positive* si toutes ses valeurs propres sont strictement positives

Si X>0 est régulière alors: $X^T M X > 0$

Une matrice $M \in \mathbb{R}^{n\times n}$ est dite *définie négative* si toutes ses valeurs propres sont strictement positives

Si X>0 est régulière alors: $X^T M X < 0$

Le Hessien d'une fonction $f:\mathbb{R}\to\mathbb{R}$ de classe $\mathcal{C}^2$ est définit par: $\nabla^2 f(x) = \left(\dfrac{\partial^2 f}{\partial x_i \partial x_j}\right) \quad i,j = 1..n$

(C'est donc la Jacobienne du gradient).

Soit une fonction $f:\mathbb{R}\to\mathbb{R}$ de classe $\mathcal{C}^2$.

Soit x* un minimum ou maximum **local**.

Condition *suffisante* au second ordre:

$\nabla^2 f(x^*) > 0$: minimum local

$\nabla^2 f(x^*) < 0$: maximum local

$\nabla^2 f(x^*) = 0$: point de selle

Conditions nécessaires d'optimalité

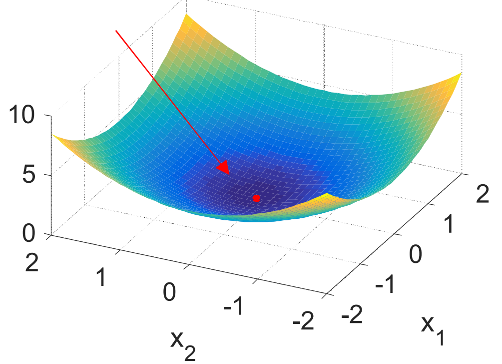
Fonction : $f(x) = a\left(x_1^{\ 2} - 0.5^2\right) + b\left(x_2^{\ 2} - 0.5^2\right)$

Gradient : $\nabla f(x) = \begin{pmatrix} 2ax_1 \\ 2bx_2 \end{pmatrix}$ $\qquad \nabla f(x) = 0 \Leftrightarrow x = (0,0)$

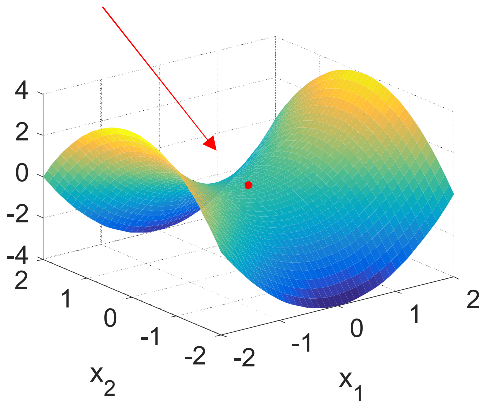
Hessien : $\nabla^2 f(x) = \begin{pmatrix} 2a & 0 \\ 0 & 2b \end{pmatrix}$ $\qquad$ Valeurs propres : $vp\left(\nabla^2 f(x)\right) = (2a, 2b)$

$a = 1, b = 1 \qquad \nabla^2 f(x) > 0$	$a = 1, b = -1 \qquad \nabla^2 f(x)\, indéfini$	$a = -1, b = -1 \qquad \nabla^2 f(x) < 0$

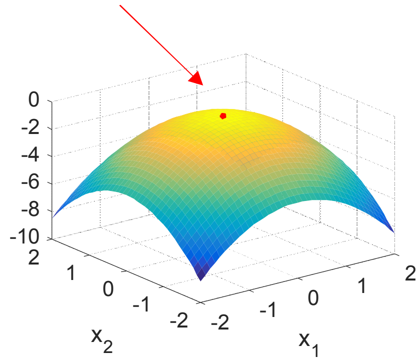
Minimum



Point de selle



Maximum



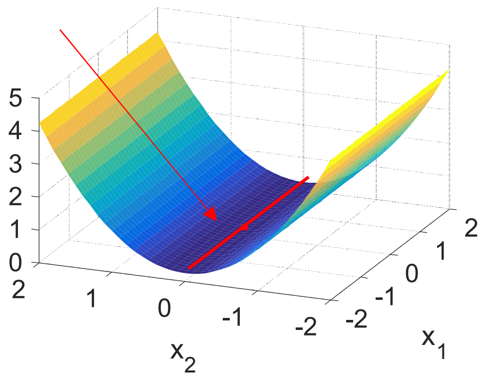
Conditions nécessaires d'optimalité

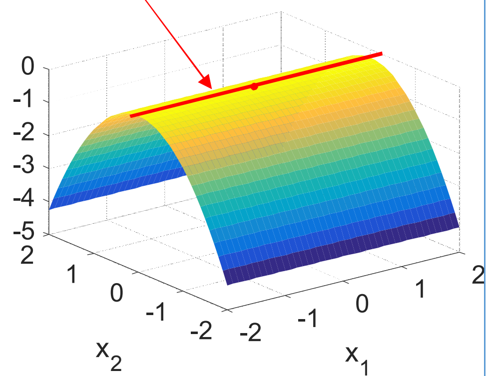
Fonction : $f(x) = a\left(x_1^2 - 0.5^2\right) + b\left(x_2^2 - 0.5^2\right)$

Gradient : $\nabla f(x) = \begin{pmatrix} 2ax_1 \\ 2bx_2 \end{pmatrix}$ $\quad \nabla f(x) = 0 \Leftrightarrow x = (0,0)$

Hessien : $\nabla^2 f(x) = \begin{pmatrix} 2a & 0 \\ 0 & 2b \end{pmatrix}$ $\quad$ Valeurs propres : $vp\left(\nabla^2 f(x)\right) = (2a, 2b)$

$a=0, b=1$ $\quad \nabla^2 f(x) \geq 0$	$a=0, b=-1$ $\quad \nabla^2 f(x) \leq 0$
Minimums	Maximums





Minimisation non contrainte

Problème à résoudre

Domaine non borné

Soit une fonction de classe $\mathcal{C}^1$: $f : \mathbb{R}^n \to \mathbb{R}$

On cherche : $x^* = \arg\min_{x \in \mathbb{R}^n} f(x)$

Et $f(x^*) = \min_{x \in \mathbb{R}^n} f(x)$

<u>Algorithme (naif):</u>

1) Calculer l'ensemble des points stationnaires $\Phi_{stat} = \{x \in \mathbb{R}^n \mid \nabla f(x) = 0\}$

2) Garder uniquement les minimums locaux $\Phi_{\min} = \{x \in \Phi_{stat} \mid \nabla^2 f(x) > 0\}$

2) Par énumération, trouver l'argmin de f $x^* = \arg\min_{x \in \Phi_{\min}} f(x)$

Minimisation non contrainte

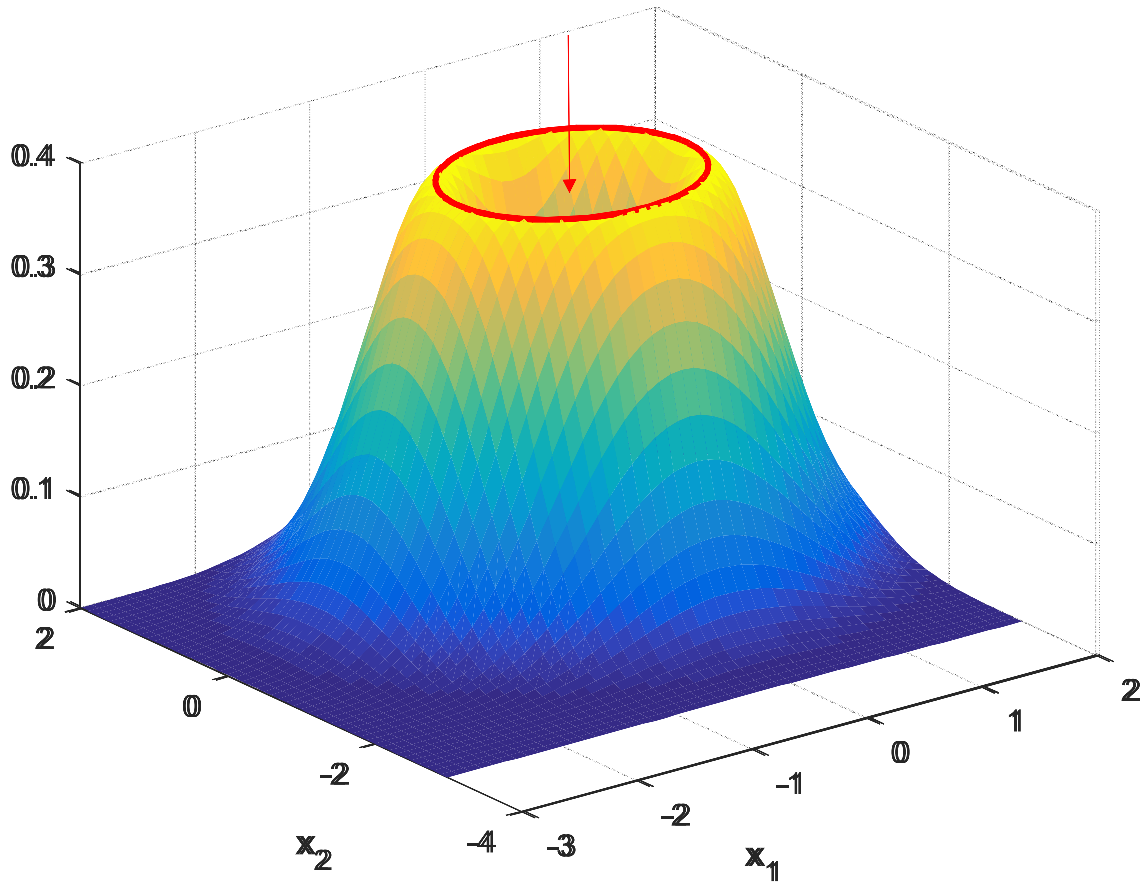
Problème à résoudre $x^* = \underset{x \in \mathbb{R}^n}{\arg\min} f(x)$

$$f(x) = \left((x_1)^2 + (x_2)^2\right) e^{-(x_1)^2-(x_2)^2}$$

$$\nabla f(x) = \begin{pmatrix} 2x_1 e^{-(x_1)^2-(x_2)^2} - 2x_1\left((x_1)^2+(x_2)^2\right)e^{-(x_1)^2-(x_2)^2} \\ 2x_2 e^{-(x_1)^2-(x_2)^2} - 2x_2\left((x_1)^2+(x_2)^2\right)e^{-(x_1)^2-(x_2)^2} \end{pmatrix}$$

$$\nabla f(x) = 0 \Leftrightarrow \begin{pmatrix} x_1 - x_1\left((x_1)^2+(x_2)^2\right) \\ x_2 - x_2\left((x_1)^2+(x_2)^2\right) \end{pmatrix} = 0$$

$$\nabla f(x) = 0 \Leftrightarrow \begin{pmatrix} x_1\left(1-(x_1)^2-(x_2)^2\right) \\ x_2\left(1-(x_1)^2-(x_2)^2\right) \end{pmatrix} = 0$$



1 Racine évidente $x = (0,0)$

Tout le cercle de centre (0,0) et de rayon 1 : $1-(x_1)^2-(x_2)^2 = 0$

Minimisation non contrainte

Gradient $\nabla f\left(x\right)=\begin{pmatrix} 2x_1\left(1-\left(x_1\right)^2-\left(x_2\right)^2\right)e^{-\left(x_1\right)^2-\left(x_2\right)^2} \\ 2x_1\left(1-\left(x_1\right)^2-\left(x_2\right)^2\right)e^{-\left(x_1\right)^2-\left(x_2\right)^2} \end{pmatrix}$

Hessien: $\nabla^2 f\left(x\right)=\begin{pmatrix} 2e^{-\left(x_1\right)^2-\left(x_2\right)^2}\left(2\left(x_1\right)^4+2\left(x_1\right)^2\left(x_2\right)^2-5\left(x_1\right)^2-\left(x_2\right)^2+1\right) & 4x_1x_2e^{-\left(x_1\right)^2-\left(x_2\right)^2}\left(\left(x_1\right)^2+\left(x_2\right)^2-2\right) \\ 4x_1x_2e^{-\left(x_1\right)^2-\left(x_2\right)^2}\left(\left(x_1\right)^2+\left(x_2\right)^2-2\right) & 2e^{-\left(x_1\right)^2-\left(x_2\right)^2}\left(2\left(x_1\right)^2\left(x_2\right)^2-\left(x_1\right)^2+2\left(x_2\right)^4-5\left(x_2\right)^2+1\right) \end{pmatrix}$

Valeurs propres : (calcul symbolique) `Lambda=simplify(eig(d2fdx2));`

$$\lambda_1=2e^{-\left(x_1\right)^2-\left(x_2\right)^2}\left(2\left(x_1\right)^4+4\left(x_1\right)^2\left(x_2\right)^2-5\left(x_1\right)^2+2\left(x_2\right)^4-5\left(x_2\right)^2+1\right)$$

$$\lambda_2=-2e^{-\left(x_1\right)^2-\left(x_2\right)^2}\left(\left(x_1\right)^2+\left(x_2\right)^2-1\right)$$

Sur le cercle $\left(x_1\right)^2+\left(x_2\right)^2=1$ λ_2 est nul => le Hessien n'est ni définit positif ni définit négatif donc le cercle n'est pas un minium de f.

En (0,0), $\lambda_1=\lambda_2=2$ => le Hessien est positif donc (0,0) est un minimum local de f.

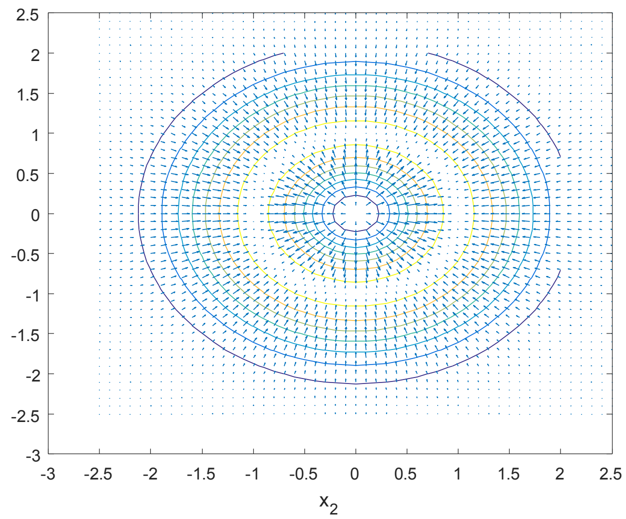
En toute rigueur, il faudrait étudier le comportement asymptotique de f pour conclure que (0,0) est un minimum *global*

Minimisation non contrainte

```
clear all
close all;
clc;
syms x1 x2

% Fonction à minimiser
f = exp(-x1^2-x2^2)*(x1^2+x2^2);
F = symfun(f,[x1 x2]);

dfdx=gradient(f,[x1,x2]);
d2fdx2= hessian(f,[x1,x2]);
d2fdx2=symfun(d2fdx2,[x1,x2]);

% Valeurs propres du Hession
lambda=simplify(eig(d2fdx2));
lambda=symfun(lambda,[x1 x2]);

% Valeurs numérique pour affichage
[X1, X2] = meshgrid(-3:.1:2,-3:.1:2);
F = double(f(X1,X2));
figure;
surf(X1,X2,F,'EdgeColor','none');
hold on;
xlabel('x_1');
ylabel('x_2');
hold on;
```

```
% Tracé du cercle
theta=0:0.1:2*pi+0.2;
xx=sin(theta);
yy=cos(theta);
zz=f(xx,yy);
plot3(xx,yy,zz,'r','linewidth',2);

% Affichage du gradient
figure;
contour(X1,X2,F);
hold on;
[X1, X2] = meshgrid(-2.5:.1:2.5,-2.5:.1:2.5);

g=gradient(f);
gg=g(X1,X2);
G1 = double(gg{1});
G2 = double(gg{2});
quiver(X1, X2, G1, G2);
xlabel('x_1');
xlabel('x_2');
```

Visualisation du gradient



Savoir-faire n°3 : Calcul symbolique

`hessian(f,[x1 x2])` : calcul du hessien de f(x1,x2)

`eig(Matrice)` : calcul des valeurs propres

`simplify(expr)` : simplification symbolique de l'expression

1) Problématiques abordées en cours
2) Ecriture d’un problème d’optimisation
3) Minimisation non contrainte
4) Minimisation avec contraintes
5) Convexité

Minimisation avec contraintes égalités

On considère le problème suivant avec contrainte égalité:

Soient les fonctions de classe $\mathcal{C}^1$: $f:\mathbb{R}^n \to \mathbb{R}$ et $h:\mathbb{R}^m \to \mathbb{R}$

Critère: $J = \min\limits_{x \in \mathbb{R}^n} f(x)$

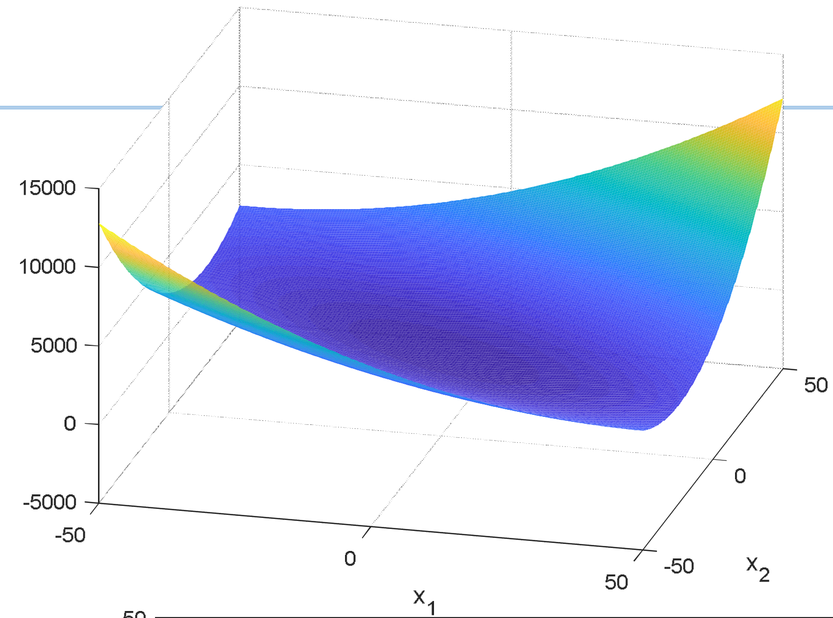
Sous la contrainte : $h(x) = 0$

=> On restreint la recherche à un sous ensemble de $\mathbb{R}^n$

Soit x^{nc} la solution du problème non contraint: $x^{nc} = \arg\min\limits_{x \in \mathbb{R}^n} f(x)$

<u>Propriété:</u> Si $h(x^{nc}) = 0$ Alors x^{nc} est aussi la solution du problème contraint

Minimisation avec contraintes égalités



<u>Rappel : résolution sans contraintes</u> $J = \min_{x \in \mathbb{R}^n} f(x)$

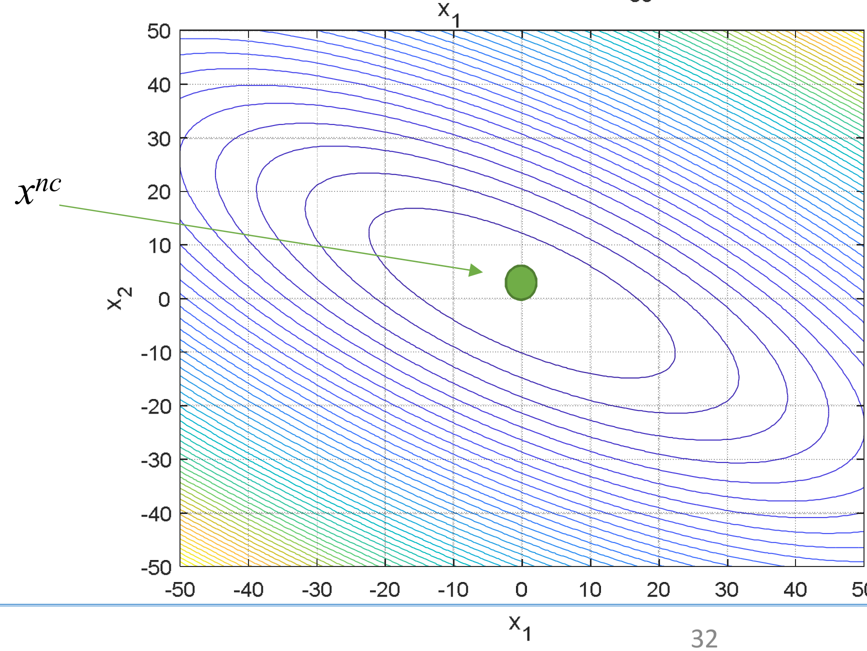
Soit $f(x) = x_1^2 + 2x_2^2 + 2x_1x_2 - 2x_1 - 4x_2$

$$\nabla f(x) = \begin{pmatrix} 2x_1 & + & 2x_2 & - & 2 \\ 2x_1 & + & 4x_2 & - & 4 \end{pmatrix} = 0$$

$$\nabla f(x) = \begin{pmatrix} 2x_1 & + & 2x_2 & - & 2 \\ 2x_1 & + & 4x_2 & - & 4 \end{pmatrix} = 0 \Leftrightarrow (x_1, x_2) = (0,1)$$

=> Un seul point stationnaire $(x_1, x_2) = (0,1)$

$$\nabla^2 f(x) = \begin{pmatrix} 2 & 2 \\ 2 & 4 \end{pmatrix} > 0$$ C'est un minimum



Minimisation avec contraintes égalités

Résolution avec une contraintes égalité: Critère: $J = \min_{x \in \mathbb{R}^n} f(x)$

Sous la contrainte : $h(x) = 0$

Idée : Définir un nouveau critère, appelé Lagrangien

$$L(x,\lambda) = f(x) + \lambda h(x)$$

$\lambda \in \mathbb{R}^n$: paramètre de Lagrange à déterminer

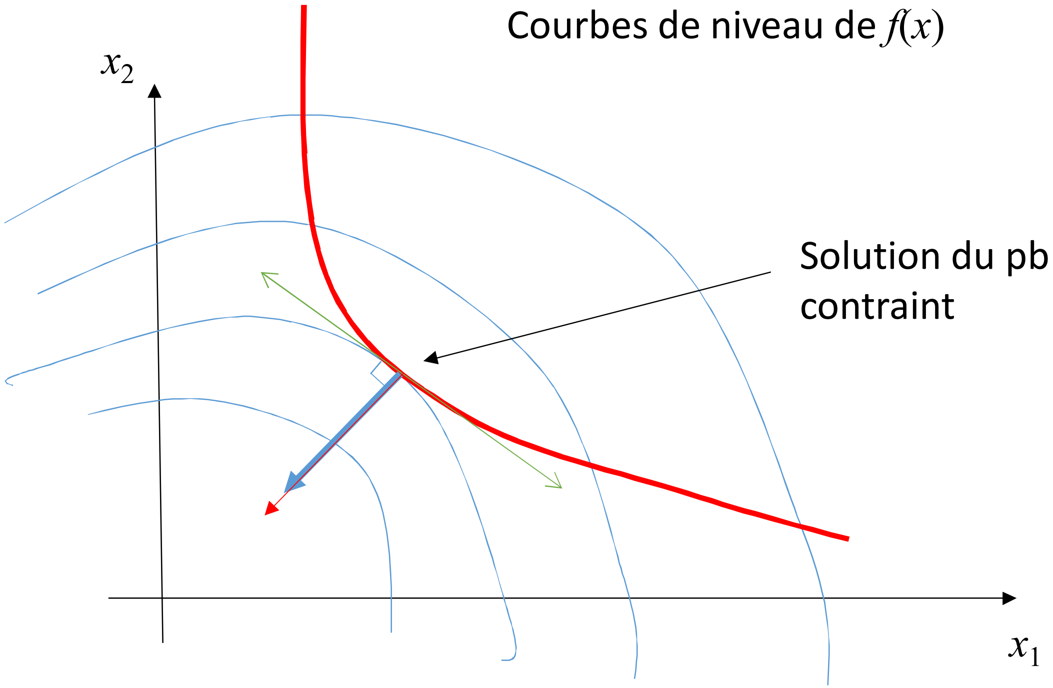
Conditions au premier ordre $\begin{cases} \nabla L(x) = 0 \\ \nabla f(x) + \lambda \nabla h(x) = 0 \end{cases}$

$$\frac{\partial L}{\partial \lambda}(x,\lambda) = 0 \Leftrightarrow h(x) = 0$$

$$\frac{\partial L}{\partial x_1}(x,\lambda) = 0 \Leftrightarrow \frac{\partial f}{\partial x_1}(x) + \lambda \frac{\partial h}{\partial x_1}(x) = 0$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\frac{\partial L}{\partial x_n}(x,\lambda) = 0 \Leftrightarrow \frac{\partial f}{\partial x_n}(x) + \lambda \frac{\partial h}{\partial x_n}(x) = 0$$



Minimisation avec contraintes égalités

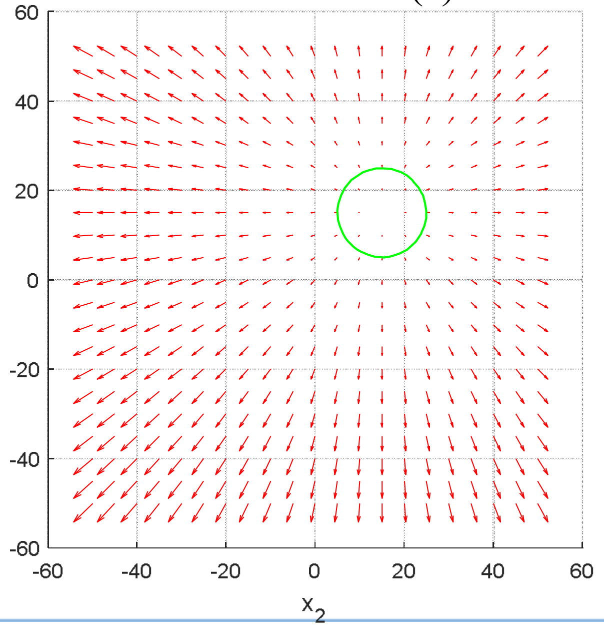
Interprétation géométrique des conditions d'optimalité au premier ordre

$$\frac{\partial L}{\partial x}(x,\lambda)=0 \Leftrightarrow \frac{\partial f}{\partial x}(x)=-\lambda\frac{\partial h}{\partial x}(x)$$

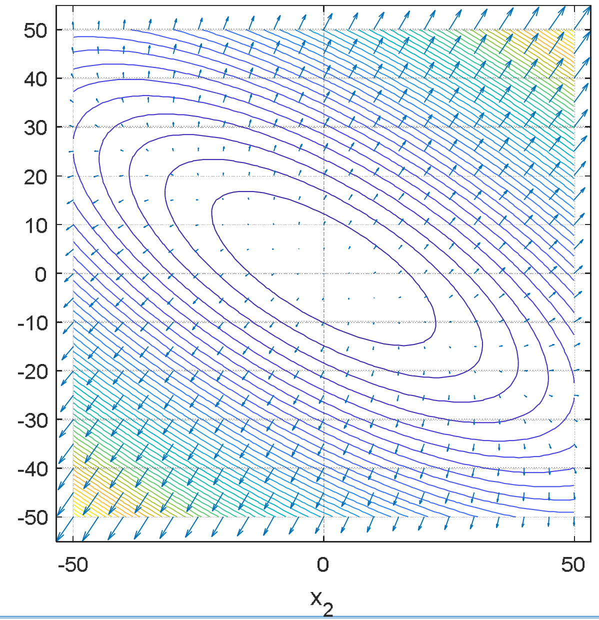
- Le gradient de la fonction et celui de la contrainte sont colinéaires sur les points stationnaires

- La courbe de niveau de la fonction et le gradient de la contrainte sont orthogonaux sur les points stationnaires

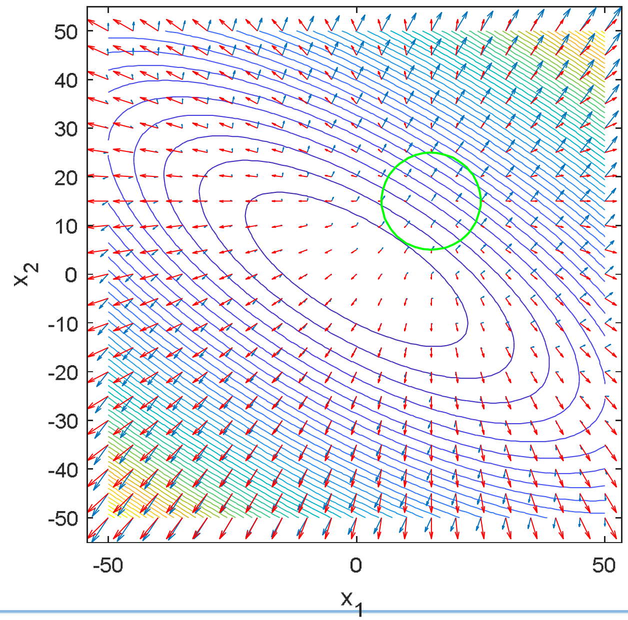
Contrainte $h(x)$



Fonction $f(x)$



Fonction $f(x)$ et contrainte $h(x)$



Minimisation avec contraintes égalités

Résolution au premier ordre:

Critère $\min_{x\in\mathbb{R}^2} f(x) = x_1^{\ 2} + 2x_2^{\ 2} + 2x_1x_2 - 2x_1 - 4x_2$

Contrainte $h(x) = (x_1 - 5)^2 + (x_2 - 5)^2 - 10^2 = 0$

Lagrangien $L(x,\lambda) = x_1^{\ 2} + 2x_2^{\ 2} + 2x_1x_2 - 2x_1 - 4x_2 + \lambda\left((x_1 - 5)^2 + (x_2 - 5)^2 - 10^2\right)$

Conditions au premier ordre pour obtenir une liste de solutions potentielles

$$\phi = \left\{(x,\lambda) \in \mathbb{R}^2 \mid \frac{\partial L(x,\lambda)}{\partial x_1} = \frac{\partial L(x,\lambda)}{\partial x_2} = \frac{\partial L(x,\lambda)}{\partial \lambda} = 0\right\}$$

Sélection du minimum

$$x^* = \underset{x\in\phi}{\operatorname{argmin}}\, f(x)$$

Minimisation avec contraintes égalités

```
clear all;
close all;
clc;

% Minimisation avec contrainte égalité
syms l x1 x2 real
Q=[1 0;1 1];
f=x1^2 + 2*x1*x2 - 2*x1 + 2*x2^2 - 4*x2;
h=(x1-15)^2+(x2-15)^2-100;

% Lagrangien
L=f+l*h;

% Recherche des points stationnaires
dL=gradient(L,[x1 x2 l]);
Stat=solve(dL,[x1 x2 l]);
```

Retourne une structure :
Stat.x1=[xxx;xxxx]
Stat.x2=[xxx;xxxx]
Stat.l=[xxx;xxxx]

```
X1potentiel=double(Stat.x1);
X2potentiel=double(Stat.x2);
Lpotentiel=double(Stat.l);
```

Crée 3 tableaux
(1 pour chaque variable)

```
f=symfun(f,[x1 x2]);
F=f(X1potentiel,X2potentiel);
[fmin,i]=min(F);
```

Calcule le minimum et l'argmin
(i: n° de la case contenant le minimum du tableau F)

```
fprintf('Solution : \n');
fprintf('  x1 : %.2e x2 : %.2e lambda : %.2\n',X1potentiel(i),X2potentiel(2),F(i));
fprintf('  paramètre de Lagrange : %.2e\n',Lpotentiel(i));
fprintf('  critere : %.2e\n',F(i));
```

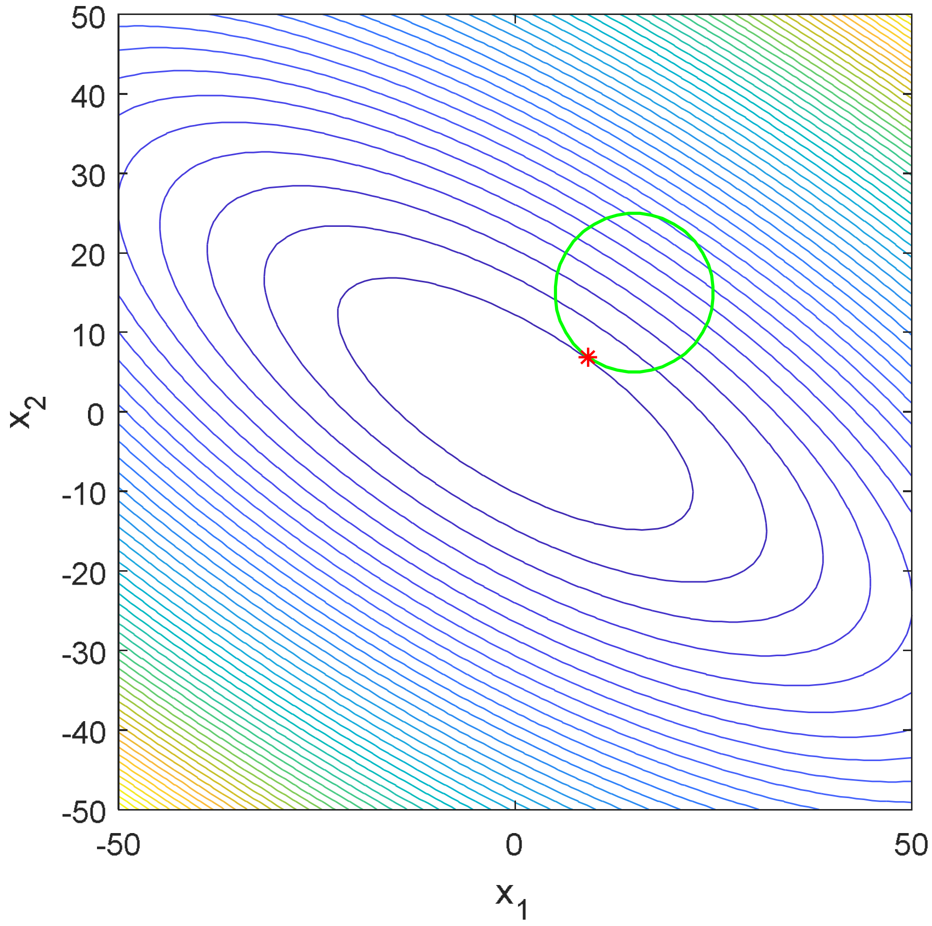
Savoir-faire n°4 : Calcul Matlab

Savoir calculer un ensemble de solution potentielles (cond. au premier ordre) et sélectionner la meilleure

`[fmin,i]=min(F)` : renvoie le minimum du tableau F et n° de la case contenant ce minimum

Minimisation avec contraintes égalités

```
h=symfun(h,[x1 x2]);
[X1, X2] = meshgrid(-50:1:50,-50:1:50);
F = double(f(X1,X2));
H= double(h(X1,X2));


figure;
contour(X1,X2,F,50);
hold on;
fc=fcontour(h,[xlim ylim],'LevelList',0);
fc.LineWidth=1;
fc.LineColor=[0 1 0];
plot(X1potentiel(i),X2potentiel(i),'r*');
xlabel('x_1');
ylabel('x_2');
axis square;
```



Minimisation avec contraintes égalités

Qualification des contraintes

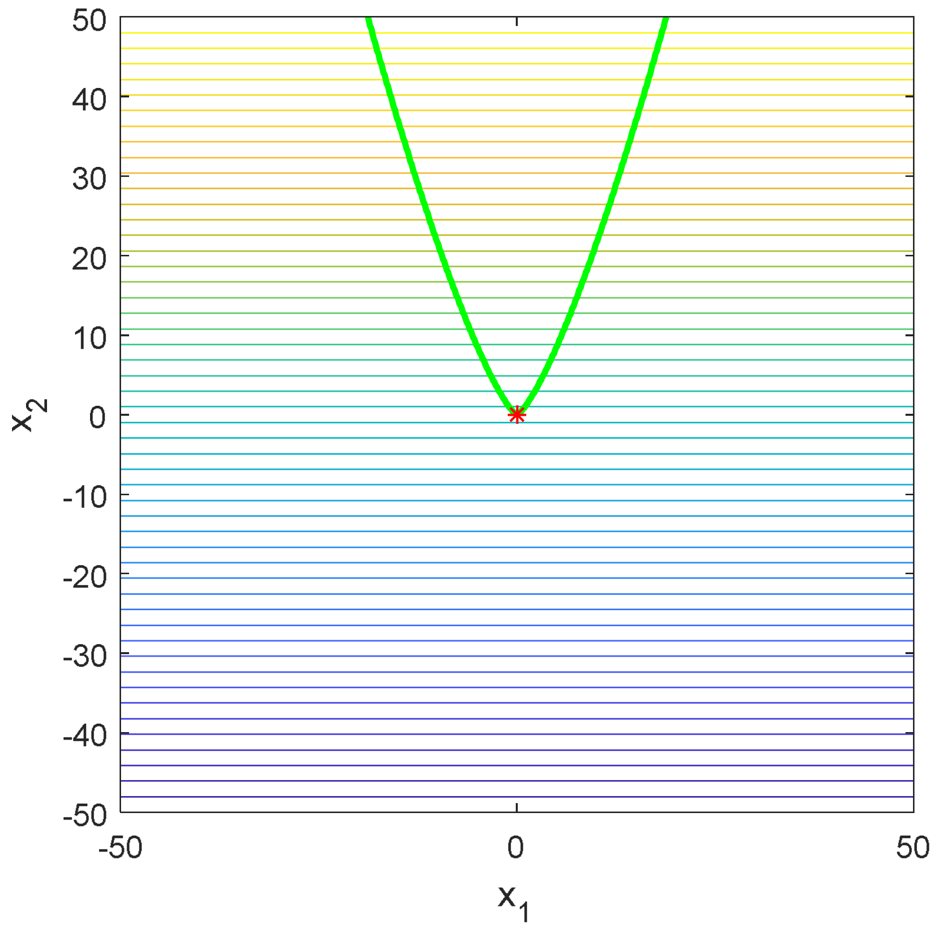
Critère $\min_{x\in\mathbb{R}^2} f(x_1,x_2) = x_2$

Contrainte $h(x) = x_2^3 - x_1^4 = 0$

Résolution « manuelle » de la contrainte: $h(x)=0 \Leftrightarrow x_2 = |x_1|^{\frac{4}{3}}$

Le problème d'optimisation réduit est: $\min_{x\in\mathbb{R}^2} f(x_1,x_2) = x_2 = |x_1|^{\frac{4}{3}}$

La solution unique est : $x_1^* = x_2^* = 0$



/exemple issu de http://www.rmi.ge/~kade/LecturesT.Kadeishvili/MathEconomics/Term4/ /

<u>Que vaut le paramètre de Lagrange pour la sol optimale ?</u>

Lagrangien $L(x_1,x_2) = x_2 + \lambda\left(x_2^3 - x_1^4\right)$

Conditions d'optimalité au premier ordre en $\left(x_1^*, x_2^*\right)$:

$\frac{L}{\partial x_1}\left(x_1^*,x_2^*\right) = 0 \Leftrightarrow -4\lambda\left(x_1^*\right)^3 = 0$ => Ne permet pas de conclure

$\frac{L}{\partial x_2}\left(x_1^*,x_2^*\right) = 0 \Leftrightarrow 1-3\lambda\left(x_1^*\right)^2 = 0$ => impossible

Minimisation avec contraintes égalités

Conditions d'optimalité au premier ordre:

Soit x* la solution du problème, et λ* le paramètre de Lagrange associé à la ou les contraintes.

1) Une seule contrainte: $\nabla f\left(x^*\right) = -\lambda^* \nabla h\left(x^*\right)$ si $\nabla h\left(x^*\right) = 0$ la condition d'optimalité n'a plus de sens

2) Plusieurs contraintes la Jacobienne des contraintes doit être de rang plein

Les contraintes doivent être indépendantes,

S'il y a *m* contraintes $h(x) = \begin{pmatrix} h_1(x) \\ \vdots \\ h_2(x) \end{pmatrix}$

$$Dh\left(x^*\right) = \begin{pmatrix} \nabla h_1\left(x^*\right) \\ \vdots \\ \nabla h_2\left(x^*\right) \end{pmatrix} = \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial h_m}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_n} \end{pmatrix}$$

Alors il faut que $rang\left(Dh\left(x^*\right)\right) = m$

Jacobienne de h

Minimisation avec contraintes égalités

Théorème: condition nécessaire d'optimalité en présence de contraintes égalité

$f : \mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^1$.

$h : \mathbb{R}^n \to \mathbb{R}^m$ de classe $\mathcal{C}^1$.

Soit x^* solution du problème suivant

$\min_{x \in \mathbb{R}^n} f(x)$ sous la contrainte $h(x) = 0$

Si $rang\left(Dh\left(x^*\right)\right) = m$ alors

Il existe $\lambda^* \in \mathbb{R}^n$ tel que les conditions suivantes sont vérifiées

$$L(x,\lambda) = f(x) + \lambda^T h(x)$$

$$\left.\begin{array}{l} \frac{\partial L}{\partial x_1}\left(x^*,\lambda^*\right) = 0 \cdots \quad \frac{\partial L}{\partial x_n}\left(x^*,\lambda^*\right) = 0 \\ \frac{\partial L}{\partial \lambda_1}\left(x^*,\lambda^*\right) = 0 \cdots \quad \frac{\partial L}{\partial \lambda_n}\left(x^*,\lambda^*\right) = 0 \end{array}\right\} \begin{array}{l} \nabla_x L(x,\lambda) = 0 \\ \nabla_\lambda L(x,\lambda) = 0 \end{array}$$

$rang\left(Dh\left(x^*\right)\right) = m$

- $\nabla h_i\left(x^*\right) \neq 0 \ \forall i = 1..m$
- Le nombre de contraintes ne peut pas excéder me nombre de variables : m≤n

Minimisation avec contraintes inégalités

On considère maintenant le problème suivant:

$f:\mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^1$.

$c:\mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^1$.

$\min_{x\in\mathbb{R}^n} f(x)$ sous la contrainte inégalité $c(x)\leq 0$

<u>Cas n°1 :</u> la solution x^* est sur la contrainte

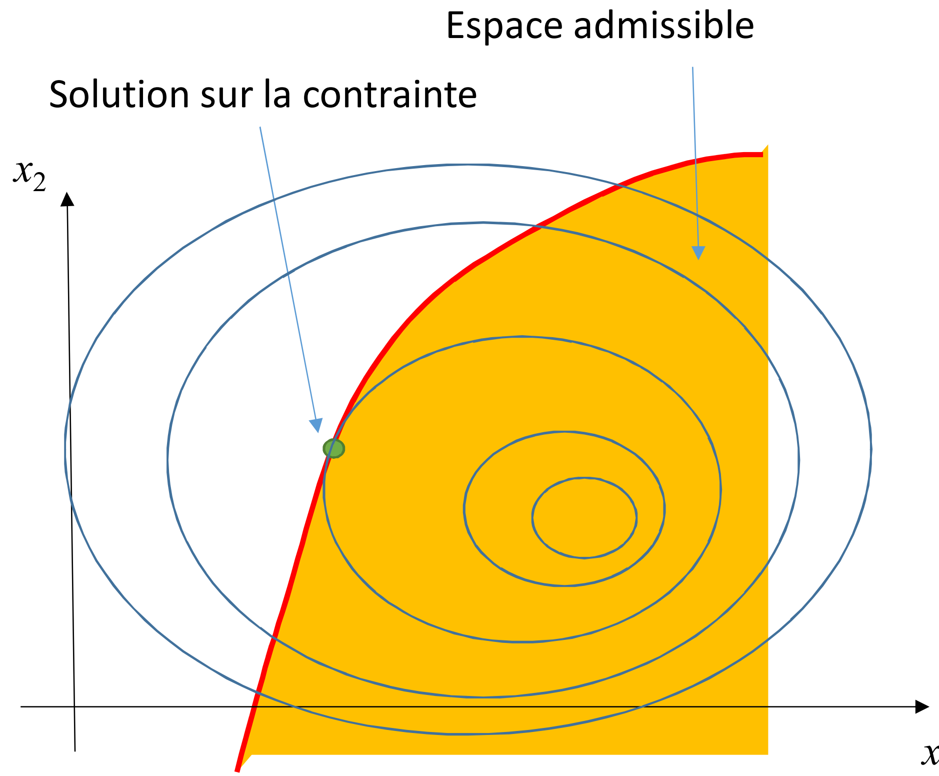
$c(x)^* = 0$

Les conditions d'optimalités au premier ordre précédentes sont vérifiées : il existe un paramètre de Lagrange $\mu^* \in \mathbb{R}^n$ tel que :

$$L(x^*,\mu^*) = f(x^*) + \mu^{*T} c(x^*)$$

$\nabla f(x^*) = -\mu^* \nabla c(x^*)$ On montre que $\mu^* \geq 0$

La condition au premier ordre $\nabla f(x^*) + \mu^* \nabla c(x^*) = 0$ est satisfaite pour $\mu^* \geq 0$.



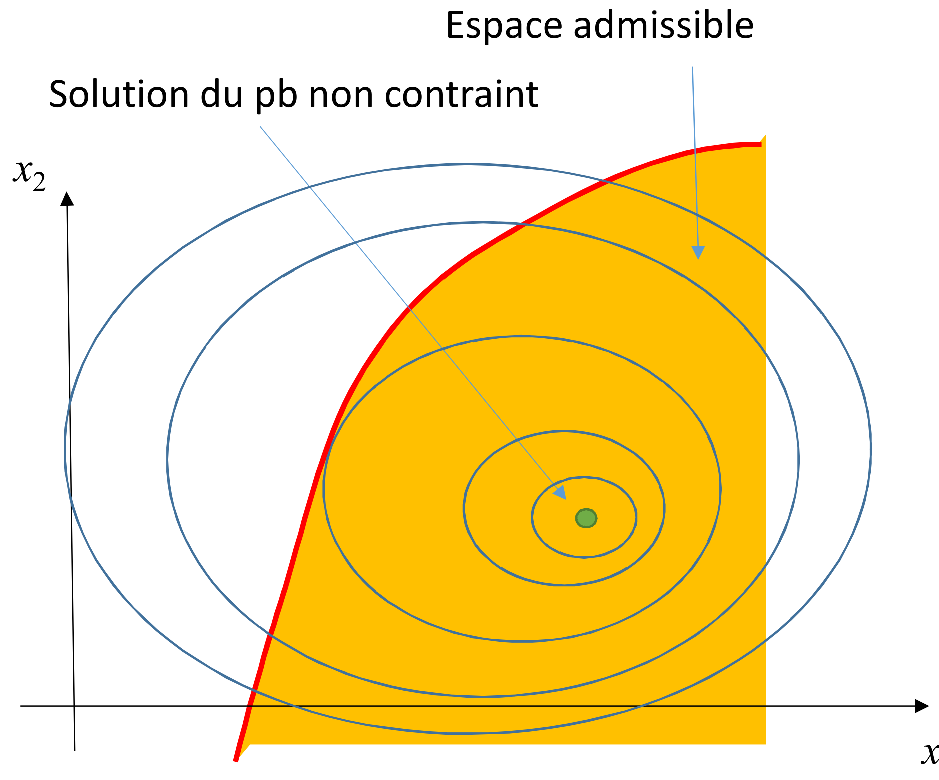
Minimisation avec contraintes inégalités

On considère maintenant le problème suivant avec une seule contrainte inégalité:

$f:\mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^1$.

$c:\mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^1$.

$\min\limits_{x\in\mathbb{R}^n} f(x)$ sous la contrainte inégalité $c(x) \leq 0$



<u>Cas n°2 :</u> la solution x^* n'est pas sur contrainte

La solution du pb non contraint est aussi solution du pb contraint

=> La contrainte n'est pas active

$$\nabla f(x^*) = 0$$

Si on pose $L(x,\mu) = f(x) + \mu^T c(x)$

La condition au premier ordre $\nabla f(x^*) + \mu^* \nabla c(x^*) = 0$

est satisfaite pour μ^*=0.

Minimisation avec contraintes inégalités

On considère maintenant le problème suivant avec une seule contrainte inégalité:

$f:\mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^1$.

$c:\mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^1$.

$\min\limits_{x\in\mathbb{R}^n} f(x)$ sous la contrainte inégalité $c(x)\leq 0$

<u>Résumé des deux cas:</u> $\mu^* \in \mathbb{R}$

Lagrangien: $L(x,\mu,\lambda) = f(x) + \mu c(x)$

Stationnarité: $\nabla L(x,\lambda) = 0 \Leftrightarrow \nabla f(x^*) + \mu_j^* \nabla c_j(x^*) = 0$

Admissibilité primale: $c(x^*) \leq 0$

Admissibilité duale: $\mu^* \geq 0$

Complémentarité: $\mu^* c(x^*) = 0$

Minimisation avec contraintes égalités & inégalités

On considère maintenant le problème suivant avec p **contraintes inégalité** :

$f:\mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^1$.

$c:\mathbb{R}^n \to \mathbb{R}^p$ de classe $\mathcal{C}^1$.

$\min\limits_{x\in\mathbb{R}^n} f(x)$ sous la contrainte inégalité $c(x)\leq 0$

De manière analogue au cas précédent, la solution (x*,μ*) vérifie :

$$\mu^* = \left(\mu^*_1,\cdots,\mu^*_p\right)^T$$

Lagrangien: $L(x,\mu,\lambda) = f(x) + \mu^T c(x)$

Stationnarité: $\nabla L(x,\lambda) = 0 \Leftrightarrow \nabla f\left(x^*\right) + \sum\limits_{j=0}^{p} \mu_j^* \nabla c_j\left(x^*\right) = 0$

Admissibilité primale: $c\left(x^*\right) \leq 0 \qquad i = 1..p$

Admissibilité duale: $\mu_i^* \geq 0 \qquad i = 1..p$

Complémentarité: $\mu_i^* c_i\left(x^*\right) = 0 \qquad i = 1..p$

Problème: la qualification des contraintes doit être étudiée uniquement pour les contraintes inégalité actives. Or elles ne sont pas connues à l'avance.

Minimisation avec contraintes égalités & inégalités

On considère maintenant le problème suivant avec $\boldsymbol{p}$ **contraintes inégalité** et $\boldsymbol{m}$ **contraintes égalité** :

$f:\mathbb{R}^n \to \mathbb{R} \qquad h:\mathbb{R}^n \to \mathbb{R}^m \qquad c:\mathbb{R}^n \to \mathbb{R}^p$ toutes de classe $\mathcal{C}^1$.

$$\min_{x\in\mathbb{R}^n} f(x)$$

sous les contraintes:

$$c(x)\leq 0 \qquad h(x)=0$$

<u>Conditions d'optimalité de Karush, Kuhn et Tucker , la solution (x^*,λ^*,μ^*) vérifie :</u>

$$\lambda^* = \left(\lambda^*{}_1,\cdots,\lambda^*{}_m\right)^T$$

$$\mu^* = \left(\mu^*{}_1,\cdots,\mu^*{}_p\right)^T$$

Lagrangien: $L(x,\mu,\lambda) = f(x) + \mu^T c(x) + \lambda^T h(x)$

Stationnarité: $\nabla L(x,\mu,\lambda) = 0 \Leftrightarrow \nabla f\left(x^*\right) + \sum_{j=0}^{p} \mu_j^* \nabla c_j\left(x^*\right) + \sum_{i=0}^{m} \lambda_i^* \nabla h_i\left(x^*\right) = 0$

Admissibilité primale: $h\left(x^*\right)=0 \qquad i=1..m$

$c\left(x^*\right)\leq 0 \qquad i=1..p$

Admissibilité duale: $\mu_i^* \geq 0 \qquad i=1..p$

Complémentarité: $\mu_i^* c_i\left(x^*\right)=0 \qquad i=1..p$

<u>Problème:</u> la qualification des contraintes doit être étudiée uniquement pour les contraintes inégalité actives. Or elles ne sont pas connues à l'avance.

=> Non traité dans ce cours

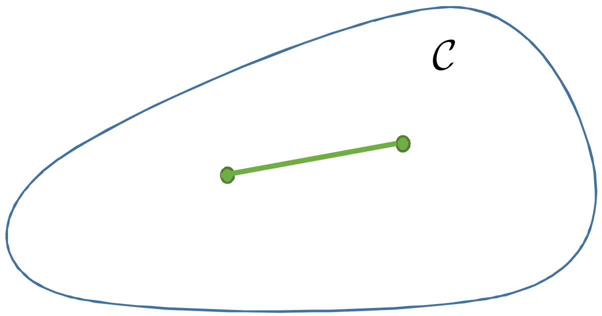
1) Problématiques abordées en cours
2) Ecriture d’un problème d’optimisation
3) Minimisation non contrainte
4) Minimisation avec contraintes
5) Convexité

Convexité du problème d'optimisation

Ensemble convexe: une ensemble $\mathcal{C}$ est dite convexe si

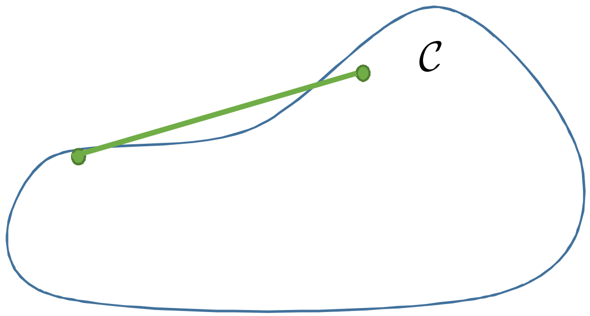
$$\forall x, y \in \mathcal{C} \qquad \forall \alpha \in [0,1] \quad \alpha x + (1-\alpha) y \in \mathcal{C}$$

Ensemble convexe



Le segment qui relie 2 points x,y de l'ensemble $\mathcal{C}$
appartient intégralement à l'ensemble $\mathcal{C}$

Ensemble non convexe

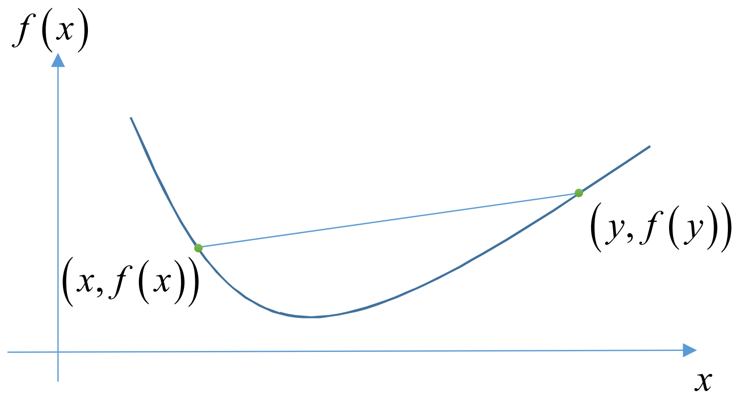


Le segment qui relie 2 points x,y de l'ensemble $\mathcal{C}$
n'appartient pas intégralement à l'ensemble $\mathcal{C}$

Convexité du problème d'optimisation

Fonction convexe: une fonction $f:\mathbb{R}^n \to \mathbb{R}$ est dite convexe si

- $dom(f)$ est un ensemble convexe
- $f(\alpha x+(1-\alpha)y)\le \alpha f(x)+(1-\alpha)f(y) \quad \forall x,y\in dom(f), 0\le\alpha\le 1$



f est convexe si la droite qui relie 2 points de son graphe est au dessus du graphe de f

Convexité du problème d'optimisation

Soit le problème $\mathcal{P}$:

$$f:\mathbb{R}^n \to \mathbb{R} \qquad h:\mathbb{R}^n \to \mathbb{R}^m \qquad c:\mathbb{R}^n \to \mathbb{R}^p$$

$$\min_{x\in\mathbb{R}^n} f(x)$$

sous les contraintes:

$$c(x) \leq 0 \quad h(x) = 0$$

Le problème $\mathcal{P}$ est dit convexe si :

- f est une fonction convexe
- L'espace des solutions admissibles $\mathcal{C}$ est convexe

$$\mathcal{C} = \left\{x \in \mathbb{R}^n \mid c(x) \leq 0; h(x) = 0\right\}$$

Le problème $\mathcal{P}$ est convexe, alors les conditions d'optimalités sont nécessaire et suffisantes.

Cf : https://see.stanford.edu/Course/EE364A pour plus d'information sur les fonctions/ensembles convexes

A quoi servent en pratique les conditions d’optimalités ?

En pratique :

- La convexité du problème d’optimisation est difficilement étudiable
- La résolution exacte des conditions d’optimalité s’avère rapidement difficile / infaisable

$$\min_{x\in\mathbb{R}} f(x) = x^5 + x^3 + 2x - 1$$

$$\frac{\partial f}{\partial x}(x) = 5x^4 + 3x^2 + 2 = 0$$ => Ne peut pas être résolu

En pratique :

- Les conditions KKT sont un très bon indicateur de la qualité d’une solution obtenue par un algorithme numérique
- On passe d’un problème contraint à un problème non contraint